

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**



FIG. 1

1/40

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): Ifa and chain (e.g. B): a

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release: DEVF9 = BPD3

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:
 Matches detected during the first 20 forward iterations

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:
 Homo sapiens Rattus norvegicus (Rat) Mus musculus (Mouse) Danio rerio (Zebra fish)

FIG. 2A

2/40

Accession	Residues	Sequence display	BPD link	WIFM link	Title	Organism	Div.	% ID	Query rgn.	Target rgn.	AJn. score	Conf.	
J	Red.Seq.View	AAF71133.1	drill through Top50BlastHits	AAF71133.1	PRO2769	Homo sapiens	PRI	13.8%	unmaskedSW	109-179	1-80	122	100% unmaskedGT
J	Red.Seq.View	CA852192.1	drill through Top50BlastHits	CA852192.1	G7c protein	Homo sapiens	PRI	9.6%	unmaskedSW	7-117	20-124	82	99.18% unmaskedGT
J	Red.Seq.View	CAA829101.1	drill through Top50BlastHits	CAA829101.1	basic transcription factor 2, 44 kD subunit	Homo sapiens	PRI	11.2%	unmaskedSW	5-164	61-225	75	98.18% unmaskedGT
J	Red.Seq.View	AAD21820.1	drill through Top50BlastHits	AAD21820.1	NC37	Homo sapiens	PRI	9.6%	unmaskedSW	7-117	318-422	82	98.88% unmaskedGT
J	Red.Seq.View	BA220761.1	drill through Top50BlastHits	BA220761.1	Not given	Homo sapiens	PRI	16.4%	unmaskedSW	5-114	1836-1950	78	98.11% unmaskedGT
J	Red.Seq.View	AAA36154.1	drill through Top50BlastHits	AAA36154.1	Not given	Homo sapiens	PRI	13.2%	unmaskedSW	5-112	10-137	73	98.16% unmaskedGT
J	Red.Seq.View	AAFD3046.1	drill through Top50BlastHits	AAFD3046.1	candidate tumor suppressor protein DICE1	Homo sapiens	PRI	13.8%	unmaskedSW	5-113	4-131	79	97.07% unmaskedGT
→	Red.Seq.View	AACT4854.1	drill through Top50BlastHits	AACT4854.1	orf, hypothetical protein	Escherichia coli	BCT	16.7%	unmaskedSW	5-99	250-335	78	95.09% unmaskedGT
J	Red.Seq.View	AA860942.1	drill through Top50BlastHits	AA860942.1	breast cancer suppressor candidate 1	Homo sapiens	PRI	16.2%	unmaskedSW	63-156	2-90	80	93.33% unmaskedGT
J	Red.Seq.View	AA467537.1	drill through Top50BlastHits	AA467537.1	glycoprotein IIIa	Homo sapiens	PRI	18.9%	unmaskedSW	3-112	111-251	72	92.08% reverse Hit unmaskedGT
J	Red.Seq.View	AA525891.1	drill through Top50BlastHits	AA525891.1	Not given	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
J	Red.Seq.View	AA871380.1	drill through Top50BlastHits	AA871380.1	platelet membrane glycoprotein IIIa beta subunit	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74	91.34% unmaskedGT reverse Hit
J	Red.Seq.View	AA435927.1	drill through Top50BlastHits	AA435927.1	Not given	Homo sapiens	PRI	19.6%	unmaskedSW	3-112	138-277	74	91.34% unmaskedGT

FIG. 2B
1) 509 hits identified by Genome Threader only:



Redundant Sequence Display

Contains:
 2 Sub-sequences.
 0 PROSITE hits.
 0 PRINTS hits.

File Edit Search Type Help

Representative code: BAA15685 | Length: 42 | Organism name: Escherichia coli | Additional annotation

Aligned sequences

Code: BAA15685 | P76236 | AAC74854

Primary database information

GenBank

protein_id: BAA15685	Details	Organism name: Escherichia coli
EC Number: Not available	Gene Name: yzdC	Description: Not given
Other links: GI: 136112 TAXON: 162		

SWISS-PROT

Code: P0235	Details	Organism name: Escherichia coli
EC Number: Not available	Gene Name: EAH	Description: HYPOTHETICAL 49.4 kDa PROTEIN INCAPABLE OF INTERGENIC REGION

Sequence

Igands

2D-PAGE

3D-PAGE

Start

Inpharmatica

6/40

FIG. 4

File Edit View Go Communicator Help

Book Forward Reload Home Search Netscape Print Security Shop

Bookmarks Location: http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi

The Sanger Centre **Pfam**
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

Results for gi|1788084|gb|AAC74854.1|

There were no matches to Pfam-A (including borderline matches) for gi|1788084|gb|AAC74854.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_39416	233	423	3.7e-103	Align

[427 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families Hypertext linked to swisspfam

Query gi|1788084|gb|AAC74854.1|/233-423 matching Pfam-B_39416

YEAM_ECOLI 233 DLRYKNYEKRDPDSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282
gi|1788084|gb|AAC74854.1| 233 DLRYKNYEKRDPDSSQAVMFCLMDVSGSMDQSTKDMAKRFYILLYLFLSR 282

YEAM_ECOLI 283 TYKNVEVVYIRDMTQAKEUDENEFFYSQETGGTIVSSALKLMDDEVVKERY 332
gi|1788084|gb|AAC74854.1| 283 TYKNVEVVYIRDMTQAKEUDENEFFYSQETGGTIVSSALKLMDDEVVKERY 332

YEAM_ECOLI 333 NPAQWNIYRAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382
gi|1788084|gb|AAC74854.1| 333 NPAQWNIYRAQASDGDNWADDSPLCHEILAKKLLPVVRYYSYIEITRRAM 382

YEAM_ECOLI 383 QTLWREYEMLQSTFDHFMQHIRDQDDIYPVFRELFHKQHA 423
gi|1788084|gb|AAC74854.1| 383 QTLWREYEMLQSTFDHFMQHIRDQDDIYPVFRELFHKQHA 423

If you think there is anything wrong with this script, please contact Pfam

100%

FIG. 5

7/40

File Edit View Go Communicator Help

Bookmarks Location <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=...>

LOCUS AAC74854 427 aa BCT 01-DEC-2000
 DEFINITION orf, hypothetical protein [Escherichia coli K12].
 ACCESSION AAC74854
 PID g1788084
 VERSION AAC74854.1 GI:1788084
 DBSOURCE locus AE000273 accession AE000273.1
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (residues 1 to 427)
 AUTHORS Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, M.A., Goeden, M.R., Rose, D.J.,
 Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 97426517
 MEDLINE 9278503
 PUBMED
 REFERENCE 2 (residues 1 to 427)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 3 (residues 1 to 427)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 REFERENCE 4 (residues 1 to 427)
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCMGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using GeneMark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 [e-mail: [mark@amber.gatech.edu">mark@amber.gatech.edu](mailto:mark@amber.gatech.edu)]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.
 FEATURES Method: conceptual translation.
 source Location/Qualifiers
 1..427
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 Protein 1..427
 /function="orf; Unknown"
 /product="orf, hypothetical protein"
 CDS 1..427
 /gene="yeaM"
 /coded_by="1788078:6385..7668"
 /transl_table=11
 /note="o427: This 427 aa ORF is 28 pct identical (43 gaps)
 to 327 residues of an approx. 312 aa protein YZDC_BACSU
 SW: P45742"
 ORIGIN 1 mtwfidrrln gknksmvnrq rflrrykaqi kqsisceaink rsvidvdsge svsiptedis
 61 epmfhggrrgg lhrhvhpgnd hfvqndrier pgggggsgs gqqqasqdgge gqdefvfqis
 121 kdeyldilfe dhalpnlkqn qqrqlyekt hragytangv panisvvrsl qnsllarrtam
 181 tackrrelha leenlaiisn sepaqlleee rlrkeiaelr akiervpfid tfdlryknye

FIG. 6A

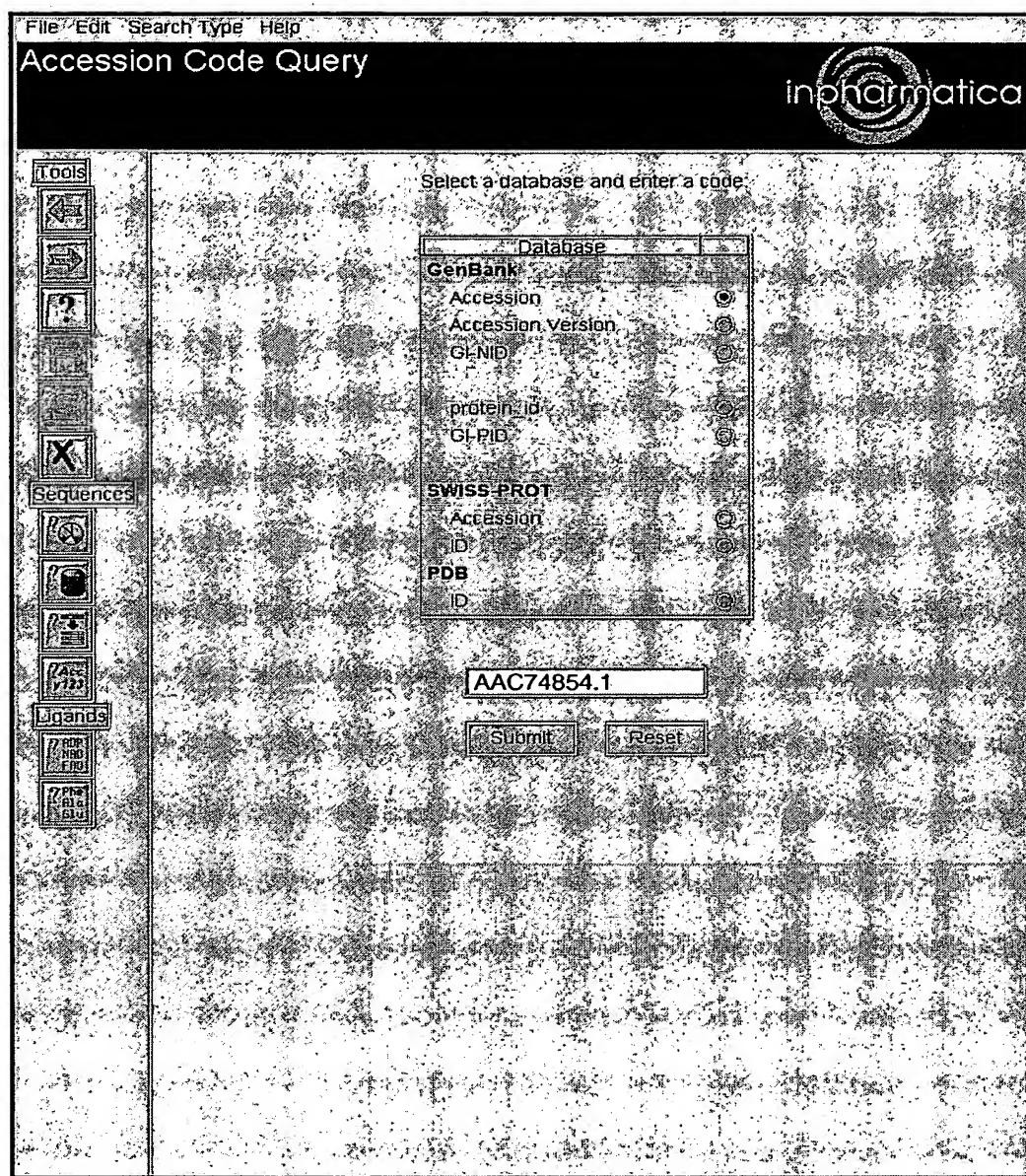


FIG. 6B Aligned Sequence Display

Query details:
AAC74854.1: orf, hypothetical protein

File Edit Search Type Help

View alignment

Inpharmatica Genome Threader results

Page: 1 of 2 Total hits: 36 Selected: 0

Cluster Detail Code Organism %ID Query/Target Amin Score Method Confidence

D 1LFAA CD11A I-DOMAIN WITH BOUND MN++ *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (85%)

D 1ZQ0A CD11A I-DOMAIN WITH BOUND MAGNESIUM *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (85%)

D 1ZQ0B CD11A I-DOMAIN WITH BOUND MAGNESIUM *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (95%)

D 1ZQ0B CD11A I-DOMAIN WITH BOUND MAGNESIUM *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (95%)

D 1ZQ0B CD11A I-DOMAIN WITH BOUND MAGNESIUM *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (95%)

D 1ZQ0A CD11A I-DOMAIN WITH BOUND MAGNESIUM *H.sapiens* 16.7 250 - 335 5 - 89 78 Local Marginal (95%)

D 1BHQ1 MAC-1 DOMAIN MAGNESIUM COMPLEX *H.sapiens* 15.1 250 - 365 5 - 128 77 Local Low (92%)

D 1IDN1 MAC-1 DOMAIN METAL FREE *H.sapiens* 15.1 250 - 365 4 - 128 76 Local Low (84%)

D 1BHQ2 MAC-1 DOMAIN CADMIUM COMPLEX *H.sapiens* 15.1 250 - 365 4 - 128 76 Local Low (92%)

D 1BHQ1 MAC-1 DOMAIN CADMIUM COMPLEX *H.sapiens* 15.1 250 - 365 4 - 128 76 Local Low (92%)

D 1BHQ2 MAC-1 DOMAIN MAGNESIUM COMPLEX *H.sapiens* 15.1 250 - 365 4 - 128 76 Local Low (92%)

D 1IDN2 MAC-1 DOMAIN METAL FREE *H.sapiens* 15.1 250 - 365 4 - 128 76 Local Low (84%)

D 1ATZB HUMAN VON WILLEBRAND FACTOR A3 DOMAIN *H.sapiens* 15.1 244 - 352 4 - 126 70 Local Low (94%)

D 1ATZA HUMAN VON WILLEBRAND FACTOR A3 DOMAIN *H.sapiens* 15.3 250 - 335 5 - 101 72 Local Low (83%)

D 1DGGQ1 INTEGRIN SUBSTRUCTURE OF THE INSERT *H.sapiens* 16.3 250 - 331 5 - 98 75 Local Low (92%)

D 1IDN1 I-DOMAIN FROM INTEGRIN CR3, MC22+ BO-*H.sapiens* 15.1 250 - 385 4 - 128 76 Local Low (84%)

D 1JL1M I-DOMAIN FROM INTEGRIN CR3, CR4, MN2+ BO-*H.sapiens* 15.1 250 - 385 4 - 128 76 Local Low (84%)

D 1ZON CD11A I-DOMAIN WITHOUT BOUND CATALYST *H.sapiens* 16.7 230 - 335 5 - 89 72 Local Low (88%)

D 1LFB AB CD11A I-DOMAIN WITH BOUND VILLEIN *H.sapiens* 15.0 250 - 333 5 - 91 66 Local Low (60%)

D 1OAKA CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 20.9 251 - 301 1 - 59 72 Local Low (72%)

D 1OAKB CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 21.2 250 - 301 1 - 68 69 Local Low (65%)

D 1AO3A A3 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 15.2 245 - 331 2 - 88 63 Local Low (53%)

D 1LFB AB CRYSTAL STRUCTURE OF THE VILLEIN *H.sapiens* 15.2 245 - 331 2 - 88 65 Local Low (53%)

D 1AO3B A3 DOMAIN OF VON WILLEBRAND FACTOR *H.sapiens* 18.1 242 - 335 1 - 105 53 Local Low (21%)

D 1AOXB 11 DOMAIN FROM INTEGRIN ALPHAI2-BETA1 *H.sapiens* 13.1 245 - 305 166 - 226 58 Local Low (21%)

D 1RPPD CRYSTAL STRUCTURE OF THE 20S PROTEASOME *Saccharomyces cerevisiae* 13.1 246 - 305 166 - 226 58 Local Low (21%)

D 1CK4B CRYSTAL STRUCTURE OF RATTUS RATTUS INTEGRIN *Rattus rattus* 10.5 250 - 331 5 - 98 55 Local Low (21%)

Icons: ? Fasta View alignment

Tools: ? Fasta View alignment

Links: GenBank SWISS-PROT PDB

File Edit Search Type Help

Aligned Sequence Display

Query details:
AAC74854.1: orf, hypothetical protein

View alignment

Total selected: 0

Reverse Maximised PSI-BLAST results:

Page 1 of 1 Total hits: 2 Selected: 0

Cluster Detail Code Title

• CAB12726.1 Not given

Conserved hypothetical protein

Organism %ID Query/ID Target/ID Bittert E-value

B. subtilis 23.0 14 - 423 2 - 385 1 - 2.00E-29

Macrocystis 15.0 34 - 371 259 - 401 -8 - 4.00E-03

Tools ? Sequences

GenBank SWISS-PROT PDB

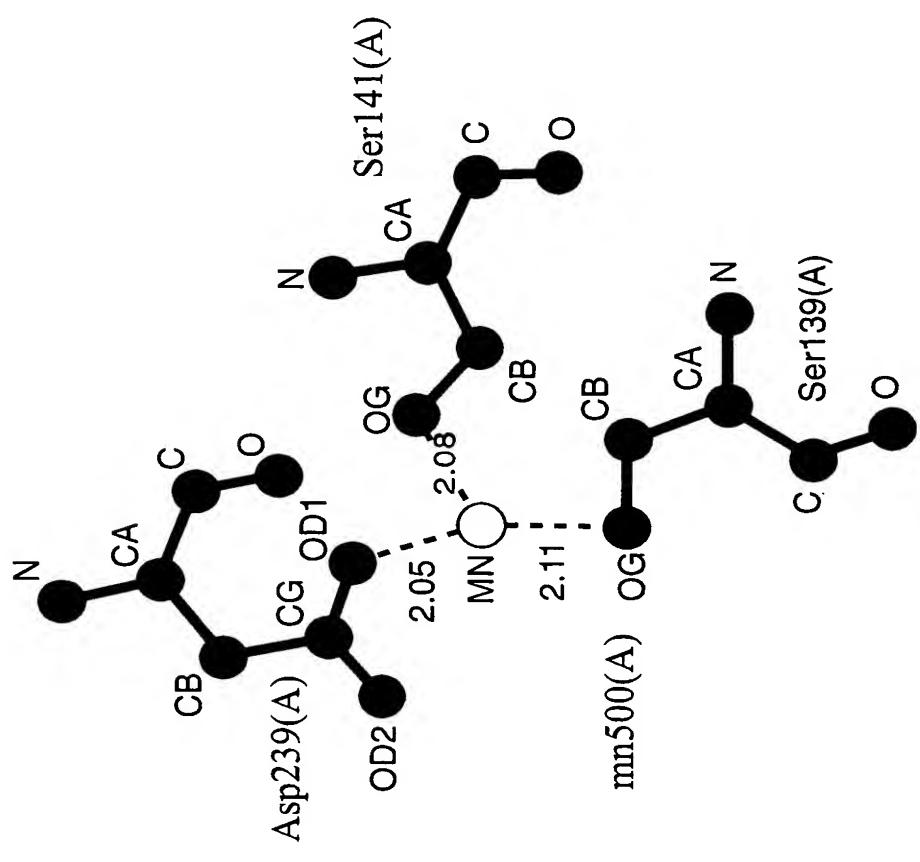
FIG. 6C

FIG. 7

AIEye output (January 4, 2002 3:07 PM)

BAA15585.1 MTWFIGDRRLNGKNKSMVNQRFLRKYKAQIKQASISEAINKRSVTDVDSSGESVSIPTEDISEPMFHQGRGGGLAHVHPGN
 1LFAA
 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 |
 DHFVQNDRIERPQQGGGSSASGQQASQDGEQGDEFVFQISKDEYLDLFFEDLALPNLKANQQRQLTEYKTHRAGYTAN
 BAA15585.1 GVPANISVVRSQNSLARRTAMTAGKRAELHAEENLAIISNSEPAQOLLEERLRAKEIAELRAKIERVPPFIOTFDLRYK
 1LFAA
 160 | 170 | 180 | 190 | 200 | 210 | 220 | 230 |
 BAA15585.1 NYEKAPDPSSQQAVMFCLMDVSGSMSMDQ9TAKDMAKRFYILLFLSRTYKNVEV
 1LFAA
 GNVDLVF
 240 | 250 | 260 | 270 | 280 | 290 | 300 | 310 |
 BAA15585.1 FYSQETGGTIVSSALKLMDEVVKERYNPAQWNIAAQASDGDNWADDSPLCHEILAKKLLPVRYYSYIEITRAHQ
 1LFAA
 HVKHMLLNTFGAINYVATEVFREELGARPDAKVLIIITDGEATD SGNIDAAKDIIRY
 320 | 330 | 340 | 350 | 360 | 370 | 380 | 390 |
 BAA15585.1 LWREYHLOSTFDNFAMQHIRDQDDIYPVFRELFFHKONATAKG
 1LFAA
 400 | 410 | 420 | 430 |
 SOETLHKFASKPASEFVKKLDTFEKLLKDLFTE.LQKKIYVIE

FIG. 8A



11fa: MN500 Chain [Å]

FIG. 8B

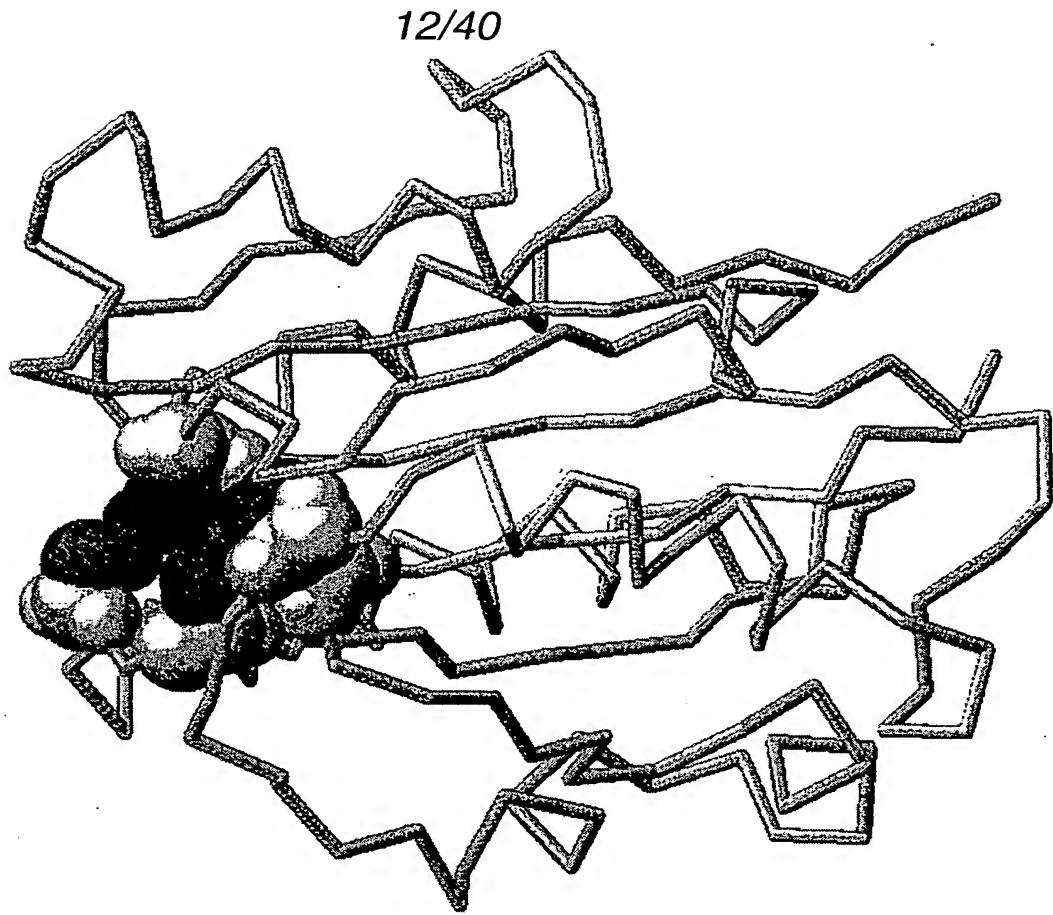
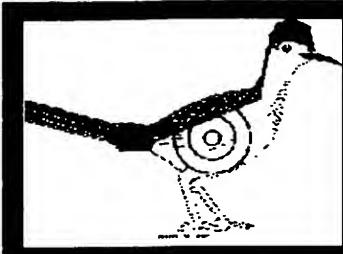


FIG. 9

13/40

File Edit View Go Communicator Help

Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface 

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

FIG. 10A

File Edit View Go Communicator Help

Bookmarks Home Search Home Escape Print Security Shop

Location: <http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBD3.pl#Mine2>

2) 82 additional hits identified by both, Genome Threader and PSI-BLAST:

Combination of Genome Threader and PSI-BLAST output: PSI-BLAST values are shown in square!

Accession	Ref ID	WWW link	Threading	Organism	Env.	% ID(GT,PSI)	Open reading frame (GT,PSI)	Target Organ. (GT,PSI)	Ah. score (GT)	Conf. (GT)	1st Iter. (PSI)	Best Iter. (PSI)	PSI-E value (PSI)		
EE6199	EE6199		INTEGRIN ALPHA-1 (LAMININ AND COLLAGEN RECEPTOR) (VLA-1) (CD46A).	Homo sapiens (Human).	PRI	51.82%	139-337, 139-337	139-337, 139-337	487	100%	1	2	2E-73		
J_	J_	dr1lthrough1050BlastHs RefSeq View	PLATELET MEMBRANE GLYCOPROTEIN IA PRECURSOR (GPIA) (COLLAGEN RECEPTOR) (INTEGRIN ALPHA-2) (VLA-2 ALPHA CHAIN) (CD46B).	Homo sapiens (Human).	PRI	89.52%	100% unmasked	2-200, 2-200	169-367, 169-367	471	100%	1	1	1E-11	
J_	J_	dr1lthrough1050BlastHs RefSeq View	Integrin alpha-11 subunit precursor	Homo sapiens	PRI	55.22%	45% unmasked	2-200, 2-200	159-355, 159-355	469	100%	1	2	4E-68	
J_	J_	dr1lthrough1050BlastHs RefSeq View	Integrin alpha 11 subunit precursor	Homo sapiens	PRI	45.22%	45% unmasked	2-200, 2-200	159-355, 159-355	469	100%	1	2	4E-68	
J_	J_	dr1lthrough1050BlastHs RefSeq View	COLLAGEN ALPHA (VIII) CHAIN PRECURSOR.	Homo sapiens (Human).	PRI	28.18%	31% unmasked	2-198, 7-199	133-318, 233-245	437	100%	1	2	2E-42	
J_	J_	dr1lthrough1050BlastHs RefSeq View	Not given	Homo sapiens	PRI	27.12%	27% unmasked	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60	
J_	J_	dr1lthrough1050BlastHs RefSeq View	Not given	Homo sapiens	PRI	22.12%	27% unmasked	2-200, 2-200	144-338, 144-338	435	100%	1	2	1E-60	
J_	J_	dr1lthrough1050BlastHs RefSeq View	d423015_1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	5-195,	5-195, 156-201	233-472, 156-201	144-338, 144-338	424	100%	2	3	5E-6	
J_	J_	dr1lthrough1050BlastHs RefSeq View	C12184_1	CARTILAGE MATRIX PROTEIN PRECURSOR (MATRILIN-1).	Homo sapiens (Human).	PRI	26.82%	26% unmasked	3-198, 3-198	271-455, 37-224	424	100%	1	2	3E-46
J_	J_	dr1lthrough1050BlastHs RefSeq View	AAAC01506_1	Type XI collagen	Homo sapiens	PRI	27.82%	30% unmasked	2-195, 2-195	133-318, 133-318	422	100%	1	2	3E-50
J_	J_	dr1lthrough1050BlastHs RefSeq View	AAAC01508_1	Integrin subunit alpha 10 precursor	Homo sapiens	PRI	46.5%	46% unmasked	2-189, 2-199	162-359, 162-359	419	100%	1	2	5E-63
J_	J_	dr1lthrough1050BlastHs RefSeq View	AAAC01502_1	matrix-4	Homo sapiens	PRI	26%	28% unmasked	1-189, 2-200	335-539, 28-223	411	100%	1	2	1E-47

FIG. 10B

		GT		Confidence			
		PRI	unmaskedSW	PRI	unmaskedGT	PRI	unmaskedGT
1	Top50BlastHits	<u>BAA91707.1</u>	Not given			142	2E-8
	Red Seq. View						
1	<u>AAA36735.1</u>						
	drill through						
1	Top50BlastHits	<u>AAA36795.1</u>	undulin 2				
	Red Seq. View						
	<u>AAD40367.1</u>						
	drill through						
1	Top50BlastHits	<u>AAD40367.1</u>	calcium-activated chloride channel-2				
	Red Seq. View						
	<u>CAA67559.1</u>						
	drill through						
1	Top50BlastHits	<u>CAA67559.1</u>	collagen VI-alpha-1 chain				
	Red Seq. View						
	<u>AAC76768.1</u>						
	drill through						
1	Top50BlastHits	<u>AAC76768.1</u>	orf, hypothetical protein				
	Red Seq. View						
	<u>CAB43000.1</u>						
	drill through						
1	Top50BlastHits	<u>CAB43000.1</u>	tellurite resistance				
	Red Seq. View						
	<u>AAA60114.1</u>						
	drill through						
1	Top50BlastHits	<u>AAA60114.1</u>	platelet membrane glycoprotein IIb				
	Red Seq. View						

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations'.

608 out of these 632 PSI-BLAST matches were identified using 'positive iterations'.

FIG. 10C

24 out of these 632 PSI-BLAST matches were identified using 'negative iterations':

bioinformatica

Redundant Sequence Display

Contains:
 2 Sub-sequences.
 0 PROSITE hits.
 0 PRINTS hits.

Representative code: P03816; Length: 427; Organism name: *Escherichia coli* [...additional annotation]

Code	Aligned sequence	Code	Aligned sequence
P03816	MAAC76768	P03816	AAAAGG0971

Primary database information

Code	Details	Organism name: <i>Escherichia coli</i>
P03818	Tax ID: 562	Gene Name: YEM
Description: HYPOTHETICAL 49.6 kDa protein in <i>Escherichia coli</i> K-12 MG1655		
Other links: ECOGENE: EGI_1730 PIR: A04400		

Sequences

GenBank	Protein ID	Details
2400	24767681	Tax ID: 562
Proteins	EC Number: Not available	Organism name: <i>Escherichia coli</i>
721	EC Number: Not available	Gene Name: YEM
Description: HYPOTHETICAL 49.6 kDa protein in <i>Escherichia coli</i> K-12 MG1655		

Start Stop

FIG. 11

FIG. 12

Pfam
Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)



Results for gi|2367274|gb|AAC76768.1|

There were no matches to Pfam-A (including borderline matches) for gi|2367274|gb|AAC76768.1|

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_15204	204	408	2.4e-108	Align

[427 residues]

.....

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|2367274|gb|AAC76768.1|204-408 matching Pfam-B_15204

```

YIEI_ECOLI 204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV 253
DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV
gi|2367274|gb|AAC76768.1| 204 DILRLLPPELATLGITELEYEFYRRLVEKQLLTYRLNGESWREKVIERPV 253

YIEI_ECOLI 254 VMKDYDEQPRGPPFIVCVDTIS6S8667WRE0CAKAPCLALMRIMLAEHRRCY 303
VMKDYDEQPRGPPFIVCVDTIS6S8667WRE0CAKAPCLALMRIMLAEHRRCY
gi|2367274|gb|AAC76768.1| 254 VMKDYDEQPRGPPFIVCVDTIS6S8667WRE0CAKAPCLALMRIMLAEHRRCY 303

YIEI_ECOLI 304 IMLFSTEIVRYELSGPQ61PQAIRFLSQQYRGGTDLASCYRAIMERLQSR 353
IMLFSTEIVRYELSGPQ61PQAIRFLSQQYRGGTDLASCYRAIMERLQSR
gi|2367274|gb|AAC76768.1| 304 IMLFSTEIVRYELSGPQ61PQAIRFLSQQYRGGTDLASCYRAIMERLQSR 353

YIEI_ECOLI 354 EWPDDRWWVISDFIRQLPDDVTSKVKEIQRVHQMRPMUAVSAMGKPGI 403
EWPDDRWWVISDFIRQLPDDVTSKVKEIQRVHQMRPMUAVSAMGKPGI
gi|2367274|gb|AAC76768.1| 354 EWPDDRWWVISDFIRQLPDDVTSKVKEIQRVHQMRPMUAVSAMGKPGI 403

YIEI_ECOLI 404 MRIRD 408
MRIRD
gi|2367274|gb|AAC76768.1| 404 MRIRD 408

```

.....

If you think there is anything wrong with this script, please contact [Pfam](#)

FIG. 13

19/40

File Edit View Go Communicator Help

Bookmarks Location: <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=protein>

LOCUS RAC76768 427 aa BCT 01-DEC-2000
 DEFINITION orf, hypothetical protein [Escherichia coli K12].
 ACCESSION RAC76768
 PID g2367274
 VERSION RAC76768.1 GI:2367274
 DBSOURCE locus RE000451 accession RE000451.1
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (residues 1 to 427)
 AUTHORS Blattner, P.R., Plumkett, S. III, Bloch, C.R., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Maphew, G.P., Gregor, J., Davis, R.W., Kirkpatrick, M.R., Goeden, M.R., Rose, D.J., Mau, B. and Shao, Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426517
 PUBMED 9278503
 REFERENCE 2 (residues 1 to 427)
 AUTHORS Blattner, P.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plumkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 REFERENCE 3 (residues 1 to 427)
 AUTHORS Blattner, P.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plumkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 REFERENCE 4 (residues 1 to 427)
 AUTHORS Plumkett, S. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined by the *E. coli* Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIMH grants M600301 and M601428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from *E. coli* K12 strain M61655. Predicted open reading frames were determined using GeneMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with *cc* site nos., unique ID nos. for the genes in the *E. coli* Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berklyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the *E. coli* Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The *E. coli* K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.
 Method: conceptual translation.
 FEATURES
 Source Location/Qualifiers
 1..427
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="M61655"
 /db_xref="taxon:83333"
 Protein 1..427
 /function="orf; Unknown"
 /product="orf, hypothetical protein"
 1..427
 /gene="yieM"
 /coded_by="complement(2367272:5249..6532)"
 /transl_table=11
 /note="f427; sequence change joins ORFs pied and piem from earlier version"
 CDS
 ORIGIN
 1 mrsrlkdav ppelteewmc yqqsqlistp qfivqlppqil dlhhrinspw aeqarqlvda
 61 nstitsalht lfliqrwlslt iqvttatnqg lleeereqlt sevgemalts gglepiladn
 121 ntaagrlwmd sagqlkrgdy qlivkygefli neqplkrila eqqlgrsreak siphndaqne
 181 tfrrtmvrep a tpeqvvdgq qsdidllrlp relatioite lenefnrrlv sklltnrh



100%



FIG. 14A

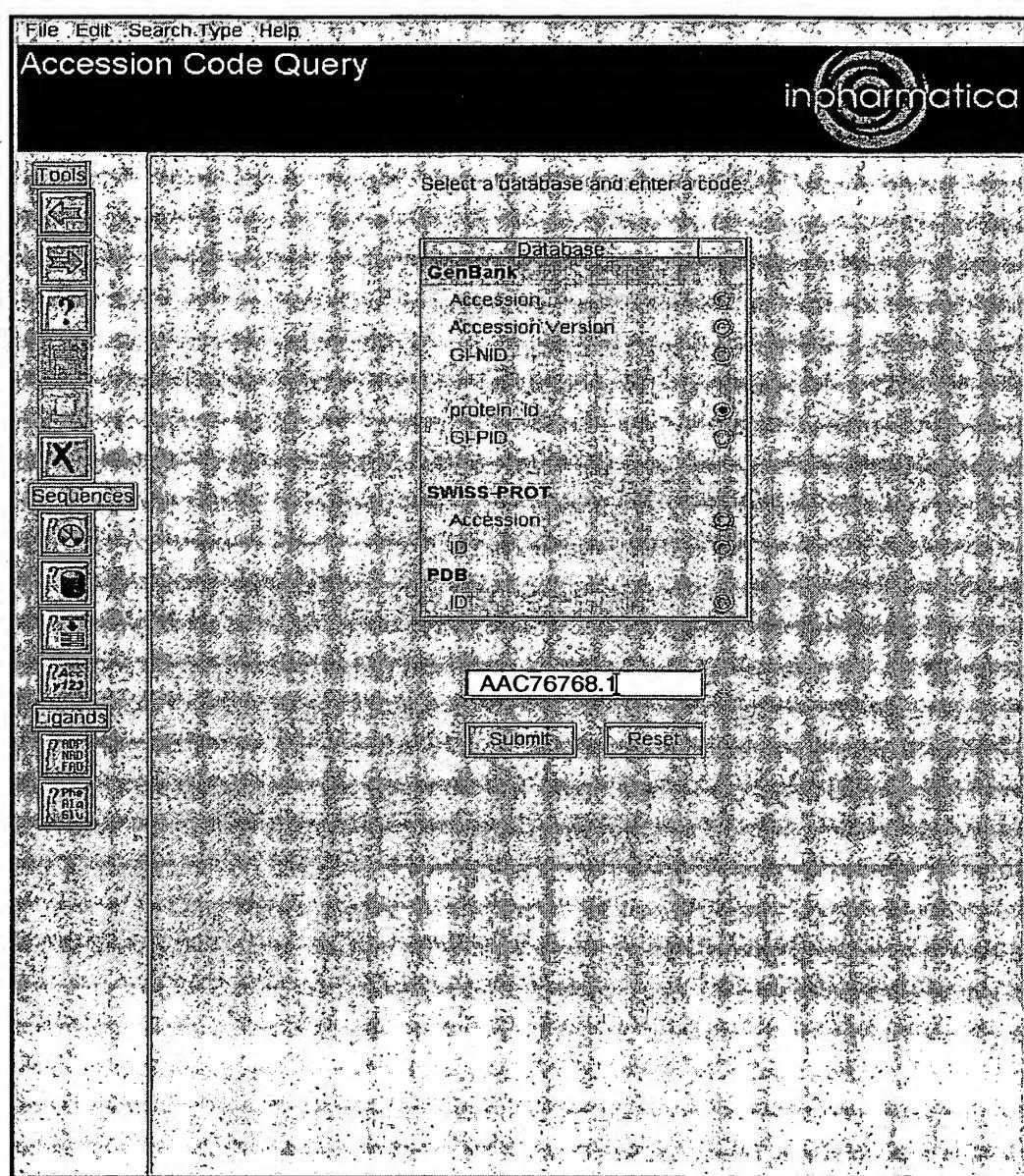


FIG. 14B Aligned Sequence Display

File Edit Search Type Help

Query details: AAC76768.1: orf, hypothetical protein

Cluster	Detail	Code	Title	Organism	SWI	COV	YRN	Targeted	AN	Score	Method	Confidence
1	ICK4B	CRYSTAL STRUCTURE OF RAT AB11 INTEGRIN BETA 1 SUBUNIT FROM INTEGRIN CRISPR	ICK4B	R. norvegicus	126	267	398	6 - 144	158	Local	Certain (100%)	
2	1JLM	I-DOMAIN FROM INTEGRIN CRISPR	1JLM	R. norvegicus	130	267	384	5 - 130	157	Local	Certain (100%)	
3	ICK4A	CRYSTAL STRUCTURE OF RAT AB11 INTEGRIN BETA 1 SUBUNIT FROM INTEGRIN CRISPR	ICK4A	R. norvegicus	126	267	398	5 - 130	156	Local	Certain (100%)	
4	1BH01	MAC-1 DOMAIN MAGNESIUM COMPLEX	1BH01	H. sapiens	188	267	384	5 - 144	149	Local	Certain (100%)	
5	1BQG2	MAC-1 DOMAIN CADMIUM COMPLEX	1BQG2	H. sapiens	188	267	384	5 - 130	148	Local	Certain (100%)	
6	1BH01	MAC-1 DOMAIN CADMIUM COMPLEX	1BH01	H. sapiens	198	267	384	5 - 130	146	Local	Certain (100%)	
7	1DQ2	MAC-1 DOMAIN METAL TREE	1DQ2	H. sapiens	198	267	384	5 - 130	146	Local	Certain (100%)	
8	1DN1	MAC-1 DOMAIN METAL TREE	1DN1	H. sapiens	188	267	384	5 - 130	146	Local	Certain (100%)	
9	1BQZ2	MAC-1 DOMAIN MAGNESIUM COMPLEX	1BQZ2	H. sapiens	188	267	384	5 - 130	146	Local	Certain (100%)	
10	1FAS9	CD11A I-DOMAIN WITH BOUND MN++	1FAS9	H. sapiens	158	264	381	1 - 124	143	Local	Certain (100%)	
11	1ZQH1	CD11A I-DOMAIN WITHOUT BOUND CAT++	1ZQH1	H. sapiens	158	264	381	1 - 124	143	Local	Certain (100%)	
12	1FAX4	CD11A I-DOMAIN WITH BOUND MN++	1FAX4	H. sapiens	146	264	354	1 - 95	141	Local	Certain (100%)	
13	1ZQH2	CD11A I-DOMAIN WITH BOUND MAGNESIUM	1ZQH2	H. sapiens	146	264	354	1 - 95	141	Local	Certain (100%)	
14	1ZQH3	CD11A I-DOMAIN WITH BOUND MAGNESIUM	1ZQH3	H. sapiens	146	264	354	1 - 95	141	Local	Certain (100%)	
15	1ZQH8	CD11A I-DOMAIN WITH BOUND MAGNESIUM	1ZQH8	H. sapiens	146	264	354	1 - 95	141	Local	Certain (100%)	
16	1D50A	NUCLEOTIDE SOLUTION STRUCTURE OF THE LISTERIA HOMINIS CD130	1D50A	H. sapiens	153	267	354	0 - 100	133	Local	Certain (100%)	
17	1ATZA	HUMAN VON WILLEBRAND FACTOR A3D1	1ATZA	H. sapiens	113	267	407	8 - 162	135	Local	Certain (100%)	
18	1WOKA	CRYSTAL STRUCTURE OF THE VON WILLEBRAND FACTOR A3D1	1WOKA	H. sapiens	131	267	398	0 - 150	111	Local	Certain (100%)	
19	1AOXA	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	1AOXA	H. sapiens	132	267	384	8 - 134	107	Local	Certain (100%)	
20	1QCS5	CRYSTAL STRUCTURE OF INTEGRIN ALPHA2-BETA1	1QCS5	H. sapiens	113	267	398	5 - 143	106	Local	Certain (100%)	
21	1QCSA	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	1QCSA	H. sapiens	113	267	398	7 - 145	102	Local	Certain (100%)	
22	1QCSB	I-DOMAIN FROM INTEGRIN ALPHA2-BETA1	1QCSB	H. sapiens	103	267	398	1 - 149	98	Local	Certain (100%)	
23	1AOXA	AS-DOMAIN OF VON WILLEBRAND FACTOR A3D1	1AOXA	H. sapiens	107	267	398	8 - 145	98	Local	Certain (100%)	
24	1AOXA	AS-DOMAIN OF VON WILLEBRAND FACTOR A3D1	1AOXA	H. sapiens	107	267	398	8 - 145	98	Local	Certain (100%)	
25	1AOXA	AS-DOMAIN OF VON WILLEBRAND FACTOR A3D1	1AOXA	H. sapiens	116	250	398	4 - 159	97	Local	Certain (100%)	
26	1ATZE	HUMAN VON WILLEBRAND FACTOR A3D1	1ATZE	H. sapiens	107	267	398	8 - 146	96	Local	Certain (100%)	
27	1ABUB	NITROGEN REGULATORY BACTERIAL PRO	1ABUB	E. coli	185	308	349	13 - 57	60	Local	Low (42%)	
28	1BQAS	ZINC FINGER PROTEIN 12 WITH UNA	1BQAS	H. sapiens	144	271	327	100 - 202	69	Local	Low (34%)	

FIG. 14C Aligned Sequence Display

Query details:
AAC76768.1: orf, hypothetical protein

File Edit Search Type Help

inparalica

22/40

Total selected 0

View alignment

Page 1 of 17 Total hits 508 Selected 0

Reverse Maximised PSI-BLAST results

Cluster	Detail	Code	Name	Qid	Qstart	Qend	Tstart	Tend	E-value
o	D	BAA80016.1	452aa long hypothetical protein	Anemix	28.0	55 - 407	202 - 431	2	2.00E-60
o	D	Q5E3B4	HYPOTHETICAL PROTEIN M00177	<i>Allosporidium</i>	27.0	46 - 496	22 - 381	3	2.00E-51
o	D	AA800610.1	conserved hypothetical protein	<i>Mannosell</i>	23.0	147 - 410	170 - 420	3	2.00E-65
o	D	AA778860.1	conserved hypothetical protein	<i>Ameliora</i>	19.0	168 - 404	24 - 448	3	2.00E-51
o	D	Q5E874	HYPOTHETICAL 45.8 kDa PROTEIN SL0103	<i>Ssp</i>	21.0	241 - 407	17 - 180	3	4.00E-39
o	D	AA884521.1	nicotinamide nucleotide oxidoreductase subunit CII	<i>Allosporidium</i>	25.0	51 - 370	324 - 335	3	5.00E-22
o	D	CAA30478.1	Not given	<i>Mimiculus</i>	20.0	247 - 416	132 - 309	4	2.00E-44
o	D	AA845044.1	Not given	<i>Haplosporidium</i>	16.0	229 - 412	17 - 305	4	3.00E-22
o	D	AAB24021.1	leucine integrin alpha chain	<i>H. saundersii</i>	16.0	223 - 412	107 - 305	4	3.00E-42
o	D	1.1.1.1	DOMAIN FROM THE SHINKAGE LINE2-BOUND	<i>H. saundersii</i>	16.0	258 - 412	101 - 158	4	2.00E-41
o	D	1.1.0.0	1 DOMAIN FROM INTEGRIN C1S-MG2+ BOUND	<i>H. saundersii</i>	16.0	253 - 412	1 - 158	4	2.00E-41
o	D	1B01.1	1 DOMAIN IN MAGNESIUM COMPLEX	<i>H. saundersii</i>	15.0	264 - 412	2 - 158	4	3.00E-41
o	D	1.1.0N2	MAC-1 DOMAIN METAL FREE	<i>H. saundersii</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	D	1.1.0N1	MAC-1 DOMAIN METAL FREE	<i>H. saundersii</i>	16.0	264 - 412	2 - 158	4	3.00E-41
o	D	1B102	MAC-1 DOMAIN IN MAGNESIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B102.2	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.1	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.2	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.3	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.4	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.5	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.6	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.7	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.8	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.9	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.10	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.11	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.12	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.13	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.14	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.15	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.16	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.17	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.18	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.19	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.20	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.21	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.22	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.23	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.24	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.25	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.26	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.27	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.28	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.29	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.30	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.31	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.32	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.33	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.34	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.35	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.36	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.37	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.38	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.39	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.40	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.41	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.42	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.43	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.44	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.45	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.46	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.47	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.48	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.49	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.50	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.51	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.52	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.53	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.54	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.55	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.56	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.57	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.58	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.59	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.60	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.61	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.62	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.63	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.64	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.65	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.66	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.67	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.68	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.69	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.70	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.71	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.72	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.73	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.74	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.75	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.76	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.77	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.78	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.79	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.80	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.81	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.82	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.83	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.84	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.85	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41
o	D	1B101.86	MAC-1 DOMAIN CADMIUM COMPLEX	<i>H. saundersii</i>	16.0	254 - 412	2 - 158	4	3.00E-41

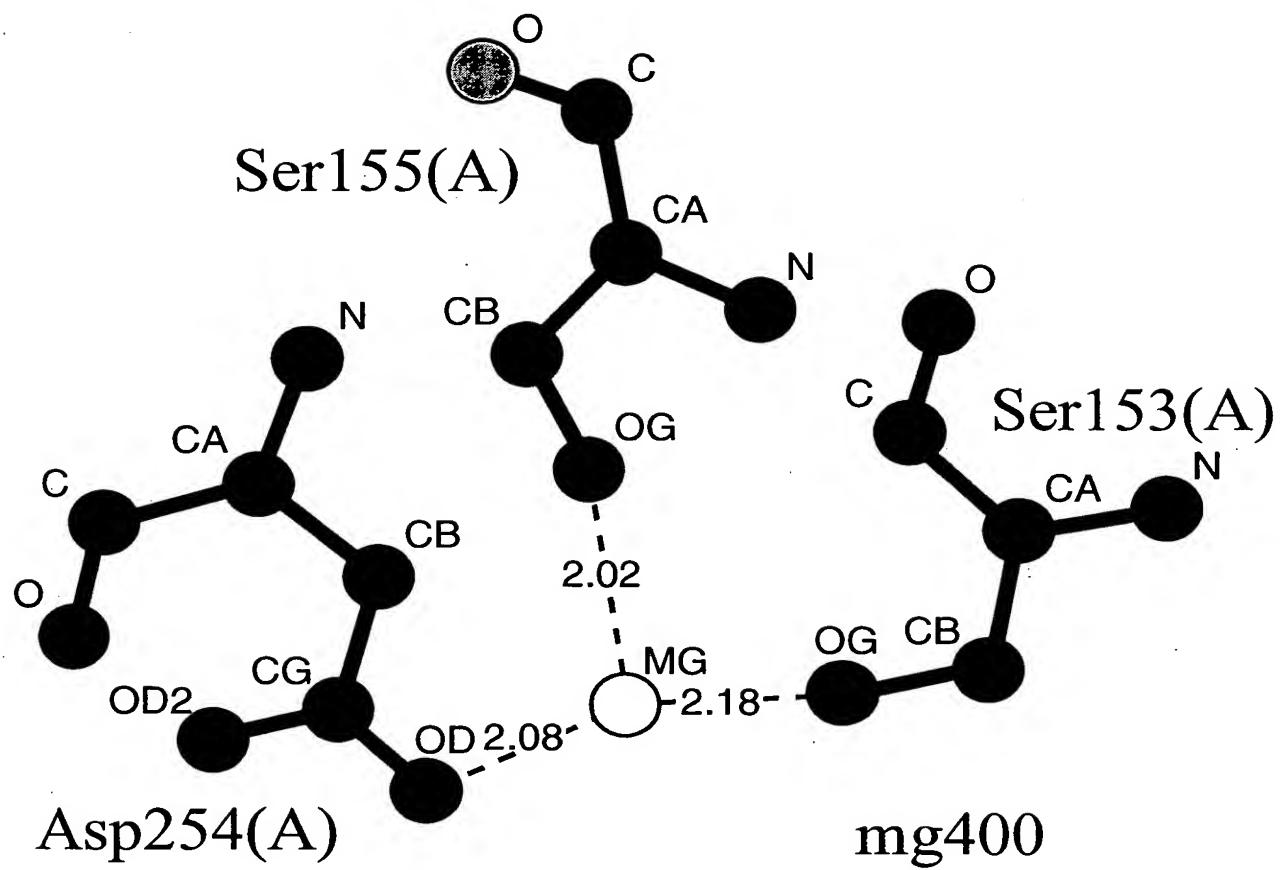
FIG. 15

AIEye output (December 13, 2000 3:07 PM)

23/40

AAC76768.1 MASPPLKDA RYPPPELTTEEVYMCYQASQALLSTPQFIVQLPQILDLHLRHLNSPWAEQARQLVDANSTITSALHTLFLQRWRW
 1AQX: A 60 90 100 110 120 130 140 150
 AAC76768.1 LSLIVQATTLNQQLLEEEPEQLLSEVQERMTLSGQLEPILADNTAAGRLYDMSAGGQLKAGDQVQLIVKYGEFLNEQP
 1AQX: A 160 170 180 190 200 210 220 230
 AAC76768.1 ELKPLAEQLGASREAKSIPRNDAQNETFARTMVAREPATYPEQVQDGLQQASDDILALLPPPELATLGITELEYEFYRPLVE
 1AQX: A 240 250 260 270 280 290 300 300
 AAC76768.1 KQLLTYPRLHGESWREKVKIERPVVHKDYDEQPRGPFIIVCVDTSGSMGGFNEQCAKAFCCLALMARI--ALAENRACYIML
 1AQX: A SCPSLID--VVVYVCDENNSIYPWDA--VKNFLEKFVQQLDIGPTKTQVGLIQ
 AAC76768.1 FSTEI--VRYELSGPQG--IEQAIIRFLSQQQFRGGTDLASCFCRAIMEALQS---RENFDADAVVVISDFIAQRLPDD
 1AQX: A YANNPRAVVFNLNTYKTKEENIVATSQTSQYGGDLTNTFGAIQYARKYAYSSAAGGPRPSATKVMVVVTDGESEH--DGSM
 AAC76768.1 VTSKVKELQRAYHQHPRFHAVAMSAHKGKPGIMRIFDH1WRFDTGMRSRLRPRWRR
 1AQX: A LKAVIDQCNHDN-ILRFGIAVGY----LNRNAUDTKNLIKEIKIAISIPTERYFFNVSDEAALLEKAG
 AAC76768.1 TLGEQIIFSIEGGT
 1AQX: A

FIG. 16A



1aox: MG400

FIG. 16B

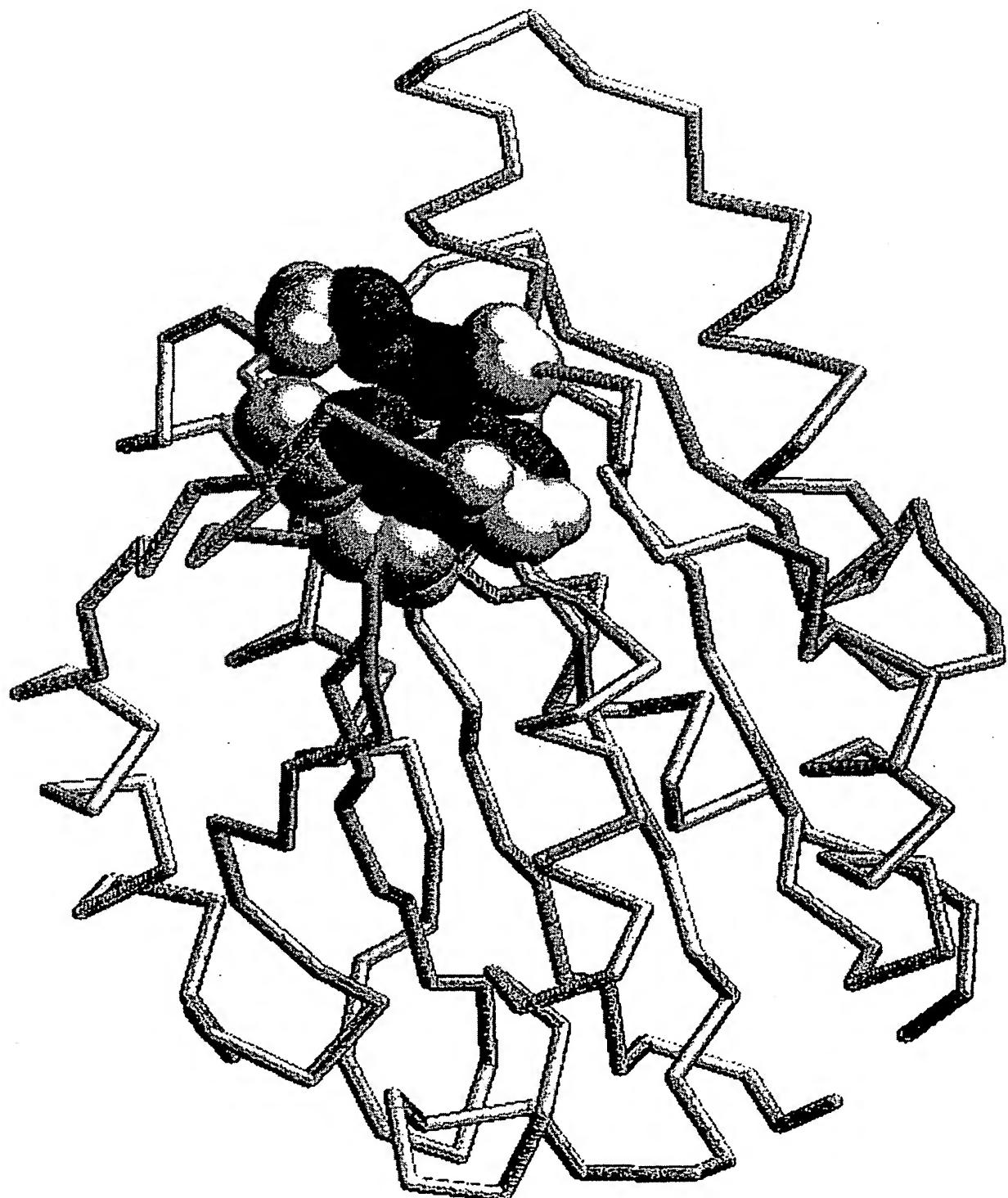
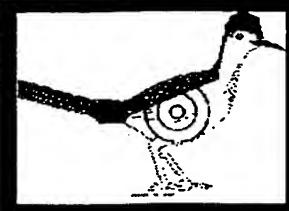


FIG. 17

26/40

File Edit View Go Communicator Help

Bookmarks Location: http://victoria.inpharmatica.co.uk/~volker/BPD3target.html

 Target Mining Interface

 inpharmatica

Select Your Query Sequence

- Enter PDB accession number (e.g. 1QMA): and chain (e.g. B):

OR

- Enter one Swiss-Prot accession (e.g. P27504) or GenBank proteinID (e.g. CAB08761.1):

Select Database

Release:

Apply Filters

- Iteration Filter: PSI-BLAST matches to be excluded:

If you select e.g. "Matches detected during the first 3 iterations" these matches will be excluded from the report (using the first_PB_iter annotation). This allows you to focus on more remote homologous which have been detected after 4 or more PSI-BLAST iterations. Matches detected using PSI-BLAST with negative iterations or using Genome-Threader are not effected by this option. However, if one match is found during the first e.g. 3 PSI-BLAST iterations and by Genome-Threader it will be excluded.

- Filter for the following SPECIES:

Homo sapiens Rattus norvegicus (Rat) Mus musculus (Mouse) Danio rerio (Zebra fish)

100% 

FIG. 18A

27/40

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location http://London-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPD3.pl

2) 81 additional hits identified by both, Genome Threader and PSI-BLAST:

Combined Genome Threader and PSI-Blast output: PSI-BLAST values are shown in maroon!

Add2list	BPD links	WWW link	Title	Organism	Div.	%ID (GT,PSI)	Query rgn. (GT,PSI)	Target rgn. (GT,PSI)	Ahn. score (GT)	Conf. (GT)	1st iter. (PSI)
↓	AAA59544.1 drill through Top50Blast1Hits	AAA59544.1	Not given	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
↓	AAB24821.1 drill through Top50Blast1Hits	AAB24821.1	leukocyte integrin alpha chain	Homo sapiens	PRI	100%, 100% unmaskedSW	1-187, 1-187	148-334, 148-334	488	100% unmaskedGT	1
↓	Q99715 drill through Top50Blast1Hits	Q99715	COLLAGEN ALPHA 1(XII) CHAIN PRECURSOR.	Homo sapiens (Human)	PRI	28.9%, 28% unmaskedSW	2-186, 2-179	439-617, 2322-2494	456	100% unmaskedGT	1
↓	AAB38702.1 drill through Top50Blast1Hits	AAB38702.1	cartilage matrix protein	Homo sapiens	PRI	28.9%, 25% unmaskedSW	2-186, 2-186	274-452, 40-221	446	100% unmaskedGT	1
↓	AAC01506.1 drill through Top50Blast1Hits	AAC01506.1	type XII collagen	Homo sapiens	PRI	28.4%, 28% unmaskedSW	2-186, 2-186	137-318, 137-318	445	100% unmaskedGT	1
↓	CAA72402.1 drill through Top50Blast1Hits	CAA72402.1	collagen type XIV	Homo sapiens	PRI	28.7%, 30% unmaskedSW	2-186, 2-186	6-185, 6-185	442	100% unmaskedGT	1
↓	AAB38547.1 drill through Top50Blast1Hits	AAB38547.1	leukointegrin alpha d chain	Homo sapiens	PRI	61%, 80% unmaskedSW	1-187, 1-187	148-334, 148-334	439	100% unmaskedGT	1
↓	CAB71222.1 drill through Top50Blast1Hits	CAB71222.1	dJ238D15.1 (collagen, type XII, alpha 1)	Homo sapiens	PRI	27.1%, 22% unmaskedSW	1-186, 2-186	293-472, 1430-1620	439	100% unmaskedGT	1
↓	CAA07569.1 drill through Top50Blast1Hits	CAA07569.1	matrin-4	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-187	344-525, 34-215	418	100% unmaskedGT	2
↓	CAB46380.1 drill through Top50Blast1Hits	CAB46380.1	dJ453C12.3 (matrin-4)	Homo sapiens	PRI	27.9%, 24% unmaskedSW	2-186, 3-186	385-566, 34-214	418	100% unmaskedGT	2

FIG. 18B

Bookmarks		Recent		Home		Logout		Security		Shop		Help	
Bookmarks	View	http://london-bridge.inpharmatica.co.uk/cgi-bin/volker/getTargetBPPD3.p	View	Home	Logout	Security	Shop	Help	Help	Help	Help	Help	Help
1	Red Star View	EntrezGene	BA001611.L	KIAA0366 protein		Homo sapiens	PRI	27.2%	Unmasked SW	73-165	390-503	66	Unmasked GI reverse HI
-	Red Star View	EntrezGene	CA021921.L	G7c protein		Homo sapiens	PRI	14.1%	Unmasked SW	6-88	20-105	76	68.6%
-	Red Star View	EntrezGene	BA015099.L	N-copine		Homo sapiens	PRI	26.0%	Unmasked SW	73-143	405-474	73	88.6%
-	Red Star View	EntrezGene	BA015099.L	INTEGRIN BETA-8 PRECURSOR.		Homo sapiens	PRI	13.1%	Unmasked SW	2-115	147-289	78	87.7%
-	Red Star View	EntrezGene	AA021820.L	NC37		Homo sapiens	PRI	14.1%	Unmasked SW	6-89	316-403	76	85.5%
-	Red Star View	EntrezGene	CA005275.L	Sec23 protein		Homo sapiens	PRI	19.7%	Unmasked SW	7-141	134-275	74	84.45%
-	Red Star View	EntrezGene	AA015099.L	copine 1		Homo sapiens	PRI	20%	Unmasked SW	73-165	384-497	67	80.79%
-	Red Star View	EntrezGene	AA015099.L	polymerase		Homo sapiens	PRI	18.1%	Unmasked SW	84-185	442-582	55	74.2%
-	Red Star View	EntrezGene	CA005280.L	4509K20.3 (Copine (similar to KIAA0636))		Homo sapiens	PRI	25%	Unmasked SW	76-143	363-429	67	88.84%
-	Red Star View	EntrezGene	CA005280.L	Not given		Homo sapiens	PRI	11.1%	Unmasked SW	6-145	373-504	56	71.48%
-	Red Star View	EntrezGene	AA015532.L	60 kDa protein (60 kDa ribonucleoprotein (RNP) (Sjogren syndrome type A antigen (SS-A)))		Homo sapiens	PRI	11.1%	Unmasked SW	6-145	373-504	58	70.11%
-	Red Star View	EntrezGene	CA015533.L	KIAA1434 protein		Homo sapiens	PRI	18.5%	Unmasked SW	22-122	74-851	75	88.76%
-	Red Star View	EntrezGene	BA002672.L	helicinase 1		Homo sapiens	PRI	20.8%	Unmasked SW	124-186	578-647	67	87.41%
-	Red Star View	EntrezGene	CA010335.L	Sec24B protein		Homo sapiens	PRI	9%	Unmasked SW	6-91	682-801	61	81.41%
-	Red Star View	EntrezGene	AA015765.L	proteolipin gamma A11 short form protein		Homo sapiens	PRI	15.6%	Unmasked SW	30-185	475-843	66	88.06%
-	Red Star View	EntrezGene	AA015766.L	proteolipin gamma A11		Homo sapiens	PRI	15.6%	Unmasked SW	30-185	475-843	66	81.1%
-	Red Star View	EntrezGene	AA015714.L	annelin 3 (annelin XXXI)		Homo sapiens	PRI	10.2%	Unmasked SW	20-184	123-308	58	82.08%
-	Red Star View	EntrezGene	CA005293.L	helicinase 1		Homo sapiens	PRI	18.5%	Unmasked SW	22-122	750-852	73	87.73%
-	Red Star View	EntrezGene	AA015915.L	Integrin beta-7 subunit		Homo sapiens	PRI	16.4%	Unmasked SW	2-55	152-204	67	88.92%
-	Red Star View	EntrezGene	AA005242.L	Integrin subunit beta 5		Homo sapiens	PRI	24.3%	Unmasked SW	2-36	130-168	73	88.92%

FIG. 18C

649 PSI-BLAST matches were identified:

849 PSI-BLAST matches were identified.

649 PSI-BLAST matches were identified.

636 *Journal of Health Politics, Policy and Law* / June 2005

322 *Journal of Business Ethics*

23 out of these 649 PSI-BLAST matches were identified using 'negative iterations'.

A2) Gamma Threadar Matches:

卷之三

File Edit View Go Communicate Help

Back Forward Home Search Netscape Stop

Bookmarks Location: ./london-bridge/inpharmatica/BP00DEV/cgi-bin/sp.pl?rep.esn=55817&password=cars_app&username=cars_app&oracle_sid=BP00DEV

Aligned annotation view for P10155 (downloading image...)

Primary database information Inpharmatica Positive matches Prints matches Text of matches Lines of matches Lines of matches

Secondary database information Inpharmatica calculated information

1: AAAA35493.1
Rep: P10155

50 100 150 200 250 300 350 400 450 500

Sequence information

Source databank	SWISSPROT
Accession code	P10155
Gene name	SSA2/CEP160
Download sequence in FASTA format	View

Links to other resources

View custom hyperlinks	View
Search in ExPASY	View
European Bioinformatics Institute	View
Online Mendelian Inheritance in Man	View
Online Mendelian Inheritance in Man	View
Protein Information Resource	View
Protein Information Resource	View

Swissprot comments:

FUNCTION	UNKNOWN
COMMENT	PROSMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT. EACH OF

Document Done

FIG. 19

31/40

FIG. 20

File Edit View Go Communicator Help

Black Forward Reload Home Search Netscape Print Security Stop

Bookmarks Location http://www.sanger.ac.uk/cgi-bin/Pfam/nph-search.cgi

The Sanger Centre **Pfam**
Protein families database of alignments and HMMs

Home | Keyword search | Protein search | DNA search | Browse Pfam | Taxonomy search | Help

Pfam

Results for gi|133251|sp|P10155|RO60_HUMAN

There were no matches to Pfam-A (including borderline matches) for gi|133251|sp|P10155|RO60_HUMAN

Matches to Pfam-B

Domain	Start	End	Evalue	Alignment
Pfam-B_8344	1	194	2.3e-103	Align
Pfam-B_10162	195	538	1.8e-165	Align

[538 residues]

Alignments of Pfam-B domains to best-matching Pfam-B sequence

Format for fetching alignments to Pfam-B families: Hypertext linked to swisspfam

Query gi|133251|sp|P10155|RO60_HUMAN/1-194 matching Pfam-B_8344

Q92787	1	MEESVNQM QPLNEKQIANSQDG YVWQVTDNMRLMRFLCF GSEGGTYYIKE	50
gi 133251 sp P10155 RO60_HUMAN	1	MEESVNQM QPLNEKQIANSQDG YVWQVTDNMRLMRFLCF GSEGGTYYIKE	50
Q92787	51	QKL GLE RAE ALIRL IEDGR GCEV I QEI KSF S QEG RTT K QEP M L P AL A I C S	100
gi 133251 sp P10155 RO60_HUMAN	51	QKL GLE RAE ALIRL IEDGR GCEV I QEI KSF S QEG RTT K QEP M L P AL A I C S	100
Q92787	101	QCS DIST K QAR FKA V SEV C R I P T H L P T I Q P K K D L K E S M K C G W G R A L R K	150
gi 133251 sp P10155 RO60_HUMAN	101	QCS DIST K QAR FKA V SEV C R I P T H L P T I Q P K K D L K E S M K C G W G R A L R K	150
Q92787	151	R I A D W Y R E K G G M A L A L R U T K Y K Q R N G W S M K D L L R L S M L K P S S E G	194
gi 133251 sp P10155 RO60_HUMAN	151	R I A D W Y R E K G G M A L A L R U T K Y K Q R N G W S M K D L L R L S M L K P S S E G	194

Query gi|133251|sp|P10155|RO60_HUMAN/195-538 matching Pfam-B_10162

008848	195	L A I V T K Y I T K G W K E V I E Y K E K A L S V E A K L L K Y L E A V E K V K R T K D D L E V	244
gi 133251 sp P10155 RO60_HUMAN	195	L A I V T K Y I T K G W K E V I E Y K E K A L S V E A K L L K Y L E A V E K V K R T K D D L E V	244
008848	245	I M L I E E M Q L V R E H L L T H M L K S K E V V K A L L Q E M P L T A L L R N L G X M T A N S V L	294
gi 133251 sp P10155 RO60_HUMAN	245	I M L I E E M Q L V R E H L L T H M L K S K E V V K A L L Q E M P L T A L L R N L G X M T A N S V L	294
008848	295	E P G N S E V S L I C E K L S N E K L L K K A R I M P F H I L I A L E T Y R A G M G L R G K L K W I	344
gi 133251 sp P10155 RO60_HUMAN	295	E P G N S E V S L I C E K L S N E K L L K K A R I M P F H I L I A L E T Y K I G M G L R G K L K W I	344
008848	345	P D K D I L Q A L D A P Y T F K I V E P T G K R F L L A V D V S A S M N Q R A L G S V L R A S T	394

FIG. 21

File Edit View Go Communicator Help

Bookmarks Location <http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=Prot>

LOCUS R060_HUMAN 538 aa **PRI** 01-FEB-1996
DEFINITION 60 KD RO PROTEIN (60 KD RIBONUCLEOPROTEIN RO) (R0RNP) (SJOGREN SYNDROME TYPE A ANTIGEN (SS-A)).

ACCESSION P10155
PID g133251
VERSION P10155 GI:133251
DBSOURCE swissprot: locus R060_HUMAN, accession P10155:
 class: standard.
 created: Mar 1, 1989.
 sequence updated: Mar 1, 1989.
 annotation updated: Feb 1, 1996.
 xrefs: gi: gi: 177782, gi: gi: 177783, gi: gi: 387656, gi: gi: 387657, gi: gi: 86722, gi: gi: 107526
 xrefs (non-sequence databases): MIM 600063, MIM 234700, PROSITE P500030

KEYWORDS Ribonucleoprotein; RNA-binding; Systemic lupus erythematosus; Antigen.
SOURCE
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Deutscher, S.L., Harley, J.B. and Keene, J.D.
TITLE Molecular analysis of the 60-kDa human Ro ribonucleoprotein
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 85 (24), 9479-9483 (1988)
MEDLINE 89071722
REMARK
REFERENCE
AUTHORS Ben-Chetrit, E., Gandy, B.J., Tan, E.M. and Sullivan, K.F.
TITLE Isolation and characterization of a cDNA clone encoding the 60-kDa component of the human SS-A/Ro ribonucleoprotein autoantigen J. Clin. Invest. 83 (4), 1284-1292 (1989)
JOURNAL 89198084
MEDLINE
REMARK
COMMENT SEQUENCE FROM N. A
 2 (residues 1 to 538)
 SEQUENCE FROM N. A
 2 (residues 1 to 538)
 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. The original entry is available from <http://www.expasy.ch/sprot> and <http://www.ebi.ac.uk/sprot>

[FUNCTION] UNKNOWN.
 [SUBUNIT] RO SMALL RIBONUCLEOPROTEINS CONSIST OF FOUR SMALL RNA MOLECULES OF 85-112 NT, EACH OF WHICH IS COMPLEXED WITH A 60 KD PROTEIN. RO RNPs MAY ALSO CONTAIN AN ADDITIONAL 52 KD PROTEIN.
 [SUBCELLULAR LOCATION] CYTOPLASMIC.
 [DISEASE] SERA FROM PATIENTS WITH SYSTEMIC LUPUS ERYTHEMATOSUS OFTEN CONTAIN ANTIBODIES THAT REACT WITH THE NORMAL CELLULAR RO PROTEIN AS IF THESE ANTIGEN WAS FOREIGN.
 [SIMILARITY] CONTAINS 1 RNA RECOGNITION MOTIF (RNP).
 [SIMILARITY] STRONG, TO XENOPUS 60 KD RO PROTEIN.

FEATURES
source Location/qualifiers
 1..538
 /organism="Homo sapiens"
 /db_xref="taxon: 9606"
 1..538
Protein 1..538
 /product="60 KD RO PROTEIN"
Region 93..98
 /region_name="Domain"
 /note="RNA-BINDING (RNP2) (BY SIMILARITY)."
Region 124..131
 /region_name="Domain"
 /note="RNA-BINDING (RNP1) (BY SIMILARITY)."
Region 239
 /region_name="Conflict"
 /note="K -> R (IN REF. 2)."
Region 515..538
 /region_name="Conflict"
 /note="GMLDMCCFDTCALDVIRNPTLDMI --> RLQNTLNLKSP (IN REF. 2)."
ORIGIN
 1 meesvngmqp lnekqiansq dgyvwqvtmdn nrlhrflcfg seggtyyike qklgleneaa
 61 lirliedgrg ceviqeqiksf sqegrttkqe pmlfalaics qcsdistkqa afkavsevcr
 121 ipthlftfq fkkdikesmk cgmwgralrk aiadwynaekg gmalalavtk ykqrngwshk
 181 dlrlrlshlkp ssegglaiytk yitkgwkevh elykekalsv etekllkyle avekvkrtkd
 241 elevihliee hrlvrehllit nhlkskeuwk allqemplta llrnlgkmta nsylepgnse
 301 vslvcekln cekllkariih pfhilialet yktghglrgk lkwrpdeeil kaldaafykt
 361 fktveptgkr fllavdvsa mnqrvlgsil nastvaaamc mvsutrtekd s yvvafsdemv
 421 pcpvttdntl qgvimamsqi paggtcdslp miwakqkntp advfivftdn etfaggvhpa
 481 iaireyrkkn dipakliveq mtsngftiad pddrgmildmc gfdtgaldvi mnftldmvi
 //

FIG. 22A

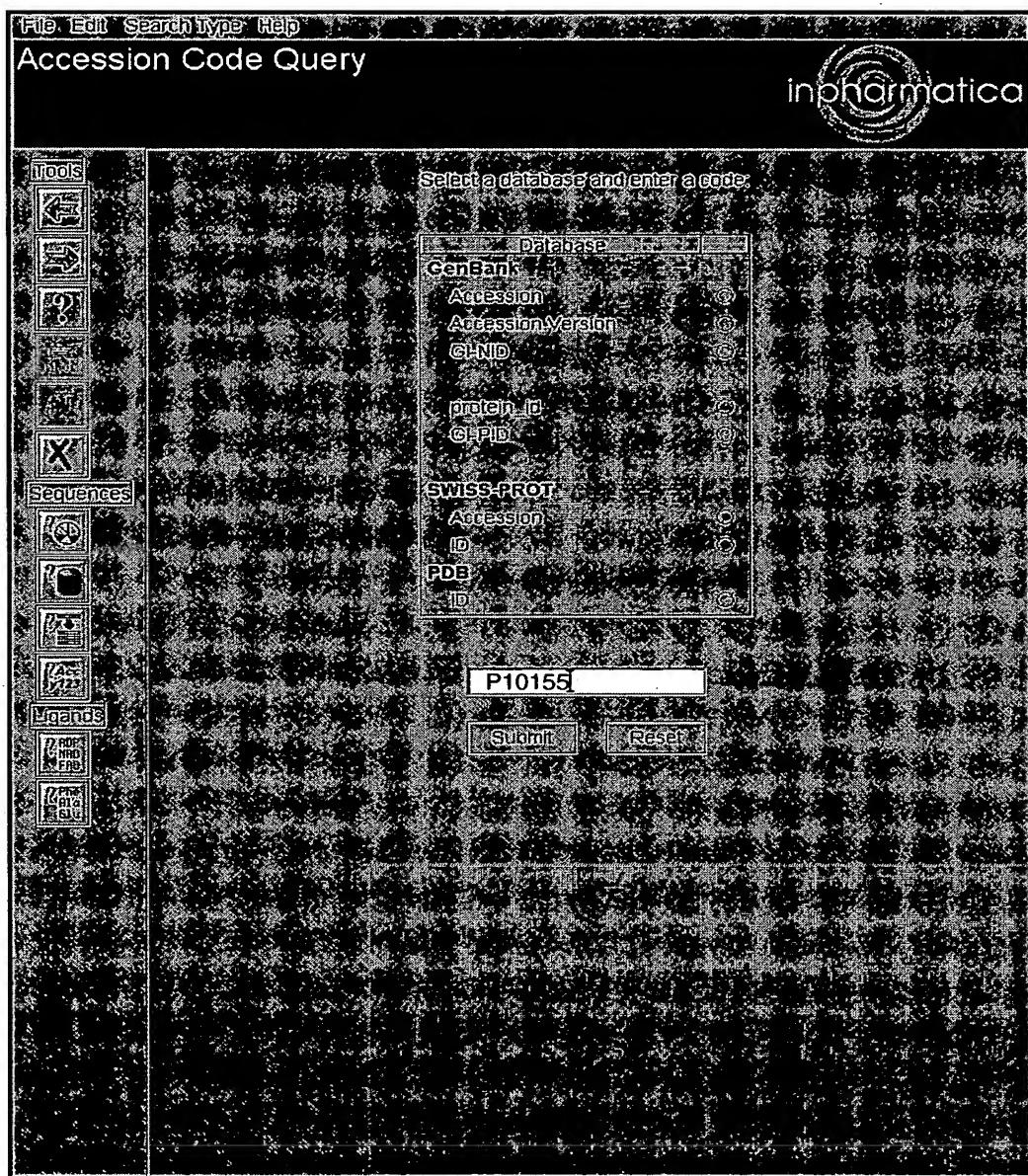


FIG. 220 Aligned Sequence Display

Query details:
P1015560

P10155: 60 kDa Ro Protein (60 kDa Ribonucleoprotein Ro) (RORNP) (Sjögren Syndrome Type A Antigen (SS-A)).

inpharmatica

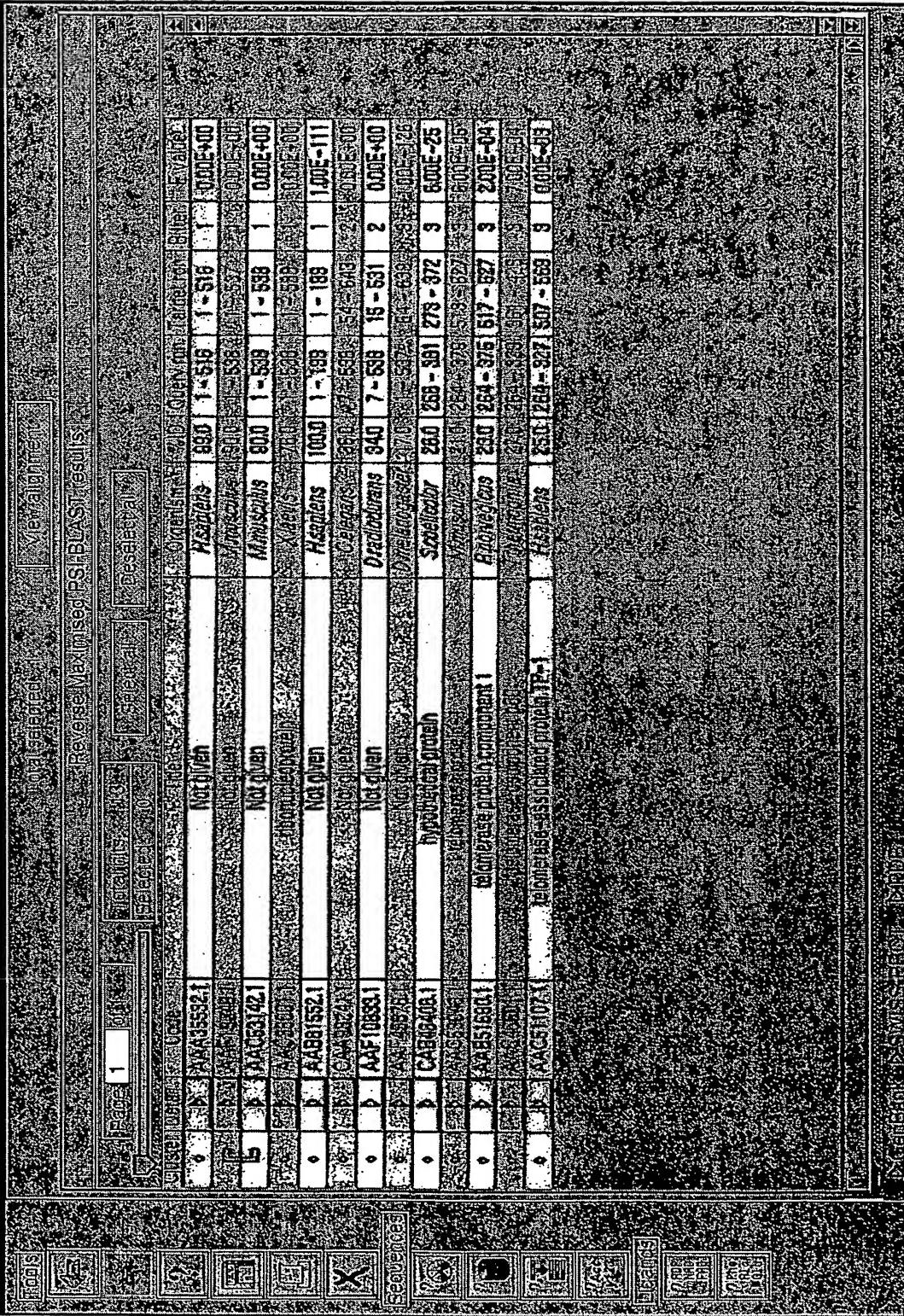


FIG. 23

AIEye output (January 2, 2001 1:29 PM)

36/40

P10155 M E E S V N Q N Q P L N E K Q I I A N S Q D G Y V W Q V T D N N P L H A F L C F G S E G G T Y Y I K E Q K L G L E N A E A L I P L I E D G R G C E V I Q E I K S F
 1JW
 P10155 S Q E G R T T K Q E P N L F A L A I C S Q C S D I S T K Q A A F K A V S E V C R I P T H L F T F I Q F K K D L K E S M K C G M W G R A L R K A I A D W Y N E K G
 1JW
 P10155 G M A L A L A V T K Y K Q R Q N G W S H K D L L R L S H L K P S S E G G L A I V T K Y I T K G W K E V H E L Y K E K A L S V E T E K L L K Y L E A V E K V K R T K D
 1JW
 P10155 E L E V I H L I E H A L Y R E H L L T N H L K S K E V V W K A L L Q E M P L T A L L R N G K W T A N S V L E P G N S E V S V L Y C E K L C N E K L L K A P I H
 1JW
 P10155 P F H I L I A L E T Y K T G H G L R G K L K W R P D E E I K A L D A A F Y K T F K T V E P T G K R F L -- L -- A V D V S A S M N N Q R V L G S I L N A S T V A
 1JW
 P10155 M E Q L K K S K T L F S L N Q Y S E E F R I H F T K E F Q N N P N P R S L V K P I T Q L L G A T H T A T G I R K V V A E L F N I T N G A R K N A F K I L V V I
 1JW
 P10155 A - - - - A N C M V V V T A T E K O S Y Y V A F S D E M V P C P V T T D M T L Q Q V L M A N S Q I - P A G G T D C S L P N I W A Q K T N T -- P A D V F I V F
 1JW
 P10155 T D N E T F A G G G V H P A I A L R E Y R K K M D I P A K L I V C G G M T S N G F - - - - - T I A D P D D O R G - N L D M C G F D T G A L D V I R N F T L D M I
 1JW
 T D G E K F G D P L G Y E D V I P E - A D R E G V I R Y Y I G V G D A F R S E K S R Q E L N T I A S K P P R D H Y F Q V N N F E - - A L K T I Q N Q L R E K I F
 1JW

P10155 1JW A

FIG. 24A

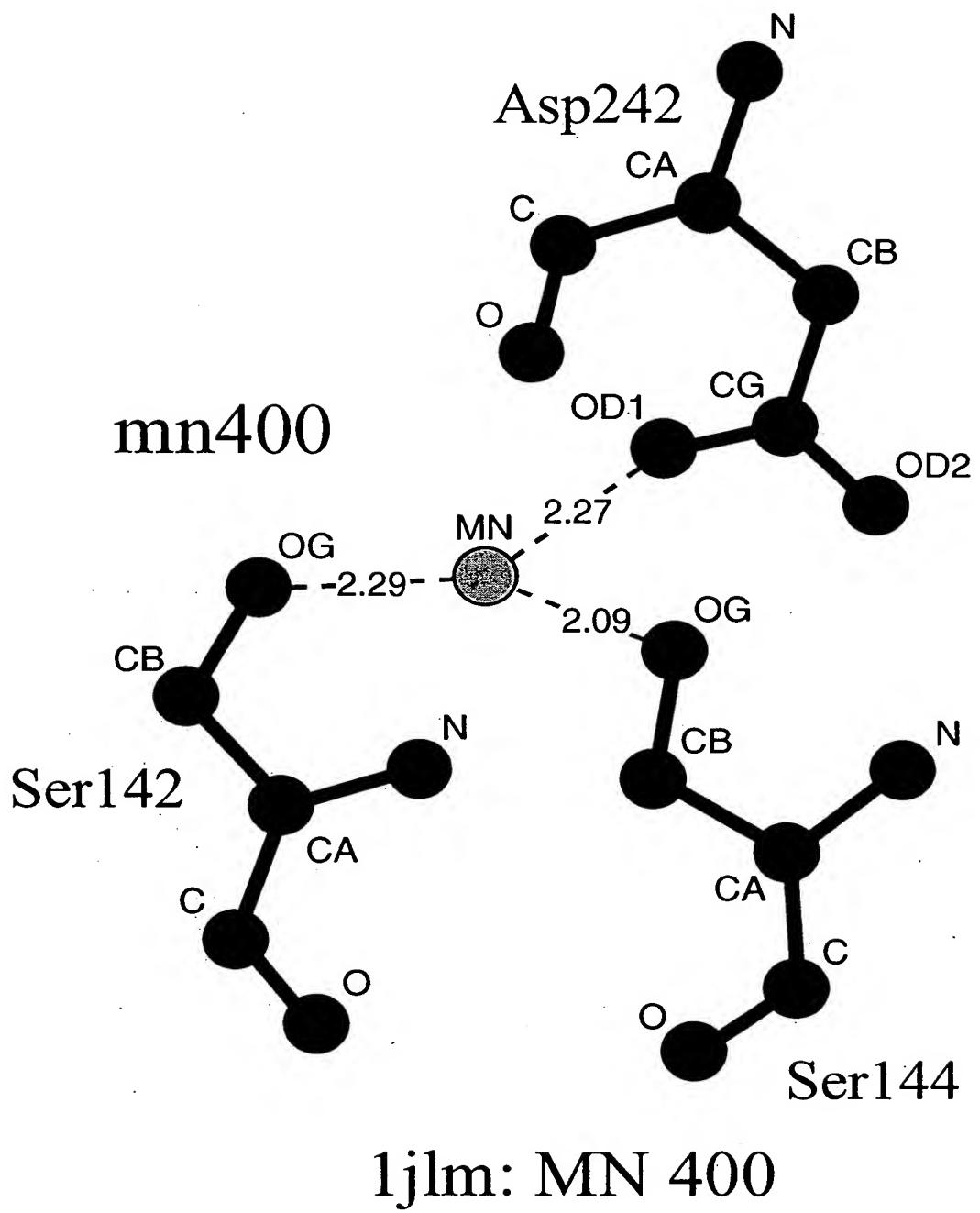


FIG. 24B

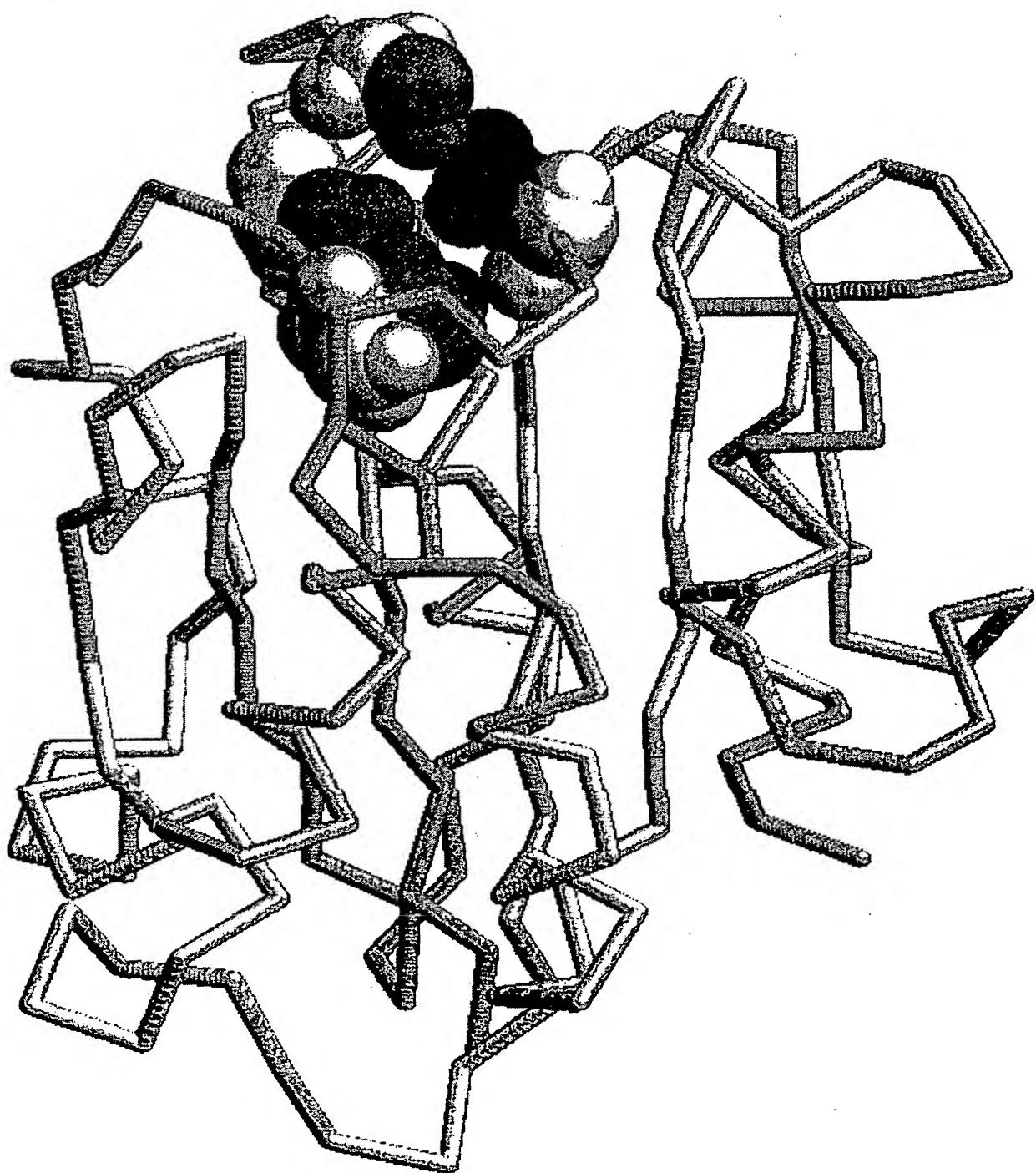


FIG. 25

AIEye output (January 4, 2001 3:18 PM)

39/40

P10155	MEESVN-----Q	10	20	30	40	50	60	70
AAF19049.1	MEGSAN-----Q							
AAC38001.1	MEATMD-----Q							
CAA98241.1	MADELNEFQEAGNFNEEALMRLSNVCA	80	90	100	110	120	130	140
P10155	RRMLESDEVITVV							
AAF19049.1	DGELKAVP							
AAC38001.1	PRQMEKVKDGGV							
CAA98241.1	ENNAGGFVFP							
P10155	VTDMNRLHRLFCFGSEGGTYIKEQKLG	150	160	170	180	190	200	210
AAF19049.1	LENAAEALIIRLIEDGRGCEV							
AAC38001.1	QEIKSFSQEGRTTKQEPMLF							
CAA98241.1	FALALICSQACSD							
P10155	ISTKQ-----I	220	230					
AAF19049.1	NTKQ-----I							
AAC38001.1	TKTKQ-----I							
CAA98241.1	HDTTKKTECPMLNAYSDYIRALHDS	240	250	260	270	280	290	300
P10155	EKGGMALAVTKYKQ	310	320	330	340	350	360	370
AAF19049.1	QRNGWSHKDLRLSH-----L							
AAC38001.1	KPSSSEGAI-----L							
CAA98241.1	KPSSSEGAI-----L							
P10155	TKDALNLA	380	390	400	410	420	430	440
AAF19049.1	MAVTKYKQ							
AAC38001.1	QRNGWSHRDLFRLAHPNLMDSRSHG							
CAA98241.1	QSEDRLEREQFLRFAVKGDLV							
P10155	TTTEKLA	450	460					
AAF19049.1	MLLT							
AAC38001.1	KYQPREGWSH-----I							
CAA98241.1	TKNEQSEEVIVAAIK							
P10155	YKEKALSV-----E	470	480	490	500	510	520	530
AAF19049.1	EEHRLVREHLLNHLKSK							
AAC38001.1	KEVVKRTKDEL							
CAA98241.1	EVKVKRTKDEL							
P10155	YKEKALSV-----E	540	550	560	570	580	590	600
AAF19049.1	EEAKLKYLEAVE							
AAC38001.1	EVKVKRTKDEL							
CAA98241.1	EVKVKRTKDEL							
P10155	YKEKALSV-----E	610	620	630	640	650	660	670
AAF19049.1	EEAKLKYLEAVE							
AAC38001.1	EVKVKRTKDEL							
CAA98241.1	EVKVKRTKDEL							
P10155	ALLRNLGKMTANSV	680	690	700	710	720	730	740
AAF19049.1	LEPGNSEVSLVC							
AAC38001.1	LEPGNSEVSLVC							
CAA98241.1	LEPGNSEVSLVC							

40/40

FIG. 25 (contd.)

AIEye output (January 4, 2001 3:18 PM)

470
480
490
500
510
520
530
540
P10155 A M S Q I P A G G T D C S L P M I W A Q K T N T P A D V F I V F T D N E T F A G G V H P A I A L R E Y R K K M D I - P A K L I V C G M T S N G F T I A D P D
AAF19049.1 A M N K V P A G N T D C S L P M I W A Q K T G T A D V F I V F T D N E T F A G Q V H P A V A L R E Y R K K M D I - P A K L I V C G M T S N G F T I A D P D
AAC38001.1 K M S D I T M G S T D C A L P M L W A Q K T N T A D I F I V F T D C E T N V E D V H P A T A L K Q Y R E K M G I - P A K L I V C A M T S N G F S I A D P D
CAA98241.1 Y V N N L D F G R T D C G L P M T W A T E N N L K F D V F I I Y T D N D T W A G E I H P F E A I K K Y R E A S G I H D A K V I V M A M Q A Y D Y S I A D P S
550
560
570
580
590
600
610
620
P10155 D R G M L D M C G F D T G A L D V I R N F T L D M I
AAF19049.1 D R G M L D M C G F D T A L D V I R N F T L D V I
AAC38001.1 D R G M L D I C G F D S G A L D V I R N F T L D L I
CAA98241.1 D A G M L D I T G F D S A V P Q I V H E F V T G K I
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
940